PA Transformation-Induced Bias: Unbiased Coefficients Do Not Imply Unbiased Quantities of Interest

Carlisle Rainey

Texas A&M University, 2010 Allen Building, College Station, TX 77843, USA. Email: crainey@tamu.edu

Abstract
Political scientists commonly focus on quantities of interest computed from model coefficients rather than on the coefficients themselves. However, the quantities of interest, such as predicted probabilities, first differences, and marginal effects, do not necessarily inherit the small-sample properties of the coefficient estimates. Indeed, unbiased coefficient estimates are neither necessary nor sufficient for unbiased estimates of the quantities of interest. I characterize this transformation-induced bias, calculate an approximation, illustrate its importance with two simulation studies, and discuss its relevance to methodological research.

Political scientists use a wide range of statistical models $y_i \sim f(\theta_i)$, where $i \in \{1, \ldots, N\}$ and f represents a probability distribution. The parameter θ_i is connected to a design matrix X of k explanatory variables and a column of ones by a link function g, so that $g(\theta_i) = X_i\beta$. In the binary logit, for example, f represents the Bernoulli probability mass function and g represents the logit function, so that $y_i \sim \text{Bernoulli}(\pi_i)$ and $\pi_i = \text{logit}^{-1}(X_i\beta)$.

The researcher usually estimates β with maximum likelihood (ML), and, depending on the choice of g and f, the estimate $\hat{\beta}$ might have desirable small-sample properties. However, ML does not produce unbiased estimates in general. For this reason, methodologists frequently use Monte Carlo simulations to assess the small-sample properties of estimators and provide users with rules of thumb about appropriate sample sizes. For example, the ML estimates of β for the binary logit are biased away from zero, leading Long (1997, p. 54) to suggest that "it is risky to use ML with samples smaller than 100, while samples larger than 500 seem adequate."

Although methodologists tend to focus on estimating model coefficients, substantive researchers tend to focus on some other quantity of interest. A quantity of interest is simply a *transformation* τ of the model coefficients. Examples include marginal effects, first and second differences, predicted probabilities and expected values, and risk ratios (King, Tomz, and Wittenberg 2000).

Fortunately, the invariance principle allows the researcher to calculate estimates of the quantities of interest from the coefficient estimates in a principled manner. The invariance principle states that if $\hat{\beta}$ is the ML estimate of β, then for any function τ, the ML estimate of $\tau(\beta)$ is $\tau(\hat{\beta})$ (King 1998, pp. 75–76, and Casella and Berger 2002, pp. 320–321). That is, researchers can simply transform the ML estimates of the model coefficients to obtain an ML estimate of the quantity of interest. Of course, if $\hat{\beta}$ is a consistent estimator of β, then $\tau(\hat{\beta})$ must be a consistent estimator of $\tau(\beta)$. But the invariance principle raises an important question: Does $\tau(\hat{\beta})$ inherit the small-sample properties of $\hat{\beta}$, such as unbiasedness or approximate unbiasedness? The answer is no; the estimates of the quantities of interest do *not* inherit the small-sample properties of the coefficient estimates. For example, a sample size of $N = 250$ that produces nearly unbiased coefficient estimates for a probit model can lead to bias in the marginal effect estimates of 25% or more.

Author's note: All computer code necessary for replication is available at https://github.com/carlislerainey/transformation-induced-bais and dx.doi.org/10.7910/DVN/CYXFB8 (Rainey 2017).

As another example, methodologists often point out the optimal small-sample properties of least squares estimators under the normal-linear model. Indeed, under the normal-linear model, least squared estimators are the best unbiased estimator. However, many substantive researchers use a simple log-transformation, so that $\log(y) \sim N(X\beta, \sigma^2)$. And while the coefficients retain their optimal properties, several substantively interesting functions of the coefficients, such as the expected value of y, the expected change in y for a given change in X, and the marginal effect of an explanatory variable on the expected value of y, do not inherit these properties.

This subtle, yet crucial, point reveals a disconnect between the work done by substantive scholars and that done by methodologists. Methodological work tends to focus on obtaining excellent estimates of the model coefficients, while substantive research tends to focus on estimating quantities of interest.

Much methodological research implicitly suggests that an approximately unbiased coefficient estimate is necessary and/or sufficient for an approximately unbiased estimate of the quantity of interest. Classically, Nagler (1994) uses Monte Carlo simulations to assess the small-sample properties of the scobit model coefficients, but he focuses on marginal effects and predicted probabilities in his illustrative application. Recently, Nieman (2015) uses simulations to assess the small-sample properties of the coefficients in his strategic probit with partial observability, but he focuses his illustrative application on the predicted probability of civil war. In order to provide more compelling tools for substantive scholars, methodologists must extend their evaluations beyond coefficient estimates to the quantities that substantive researchers typically care about.

However, the quantities of interest and likely parameter values vary dramatically across substantive applications, making it difficult or impossible for methodologists to formulate general claims about the behavior of the estimators of potential quantities of interest. Therefore, substantive researchers must not shy away from studying the behavior of their chosen estimators in a particular application, especially given their deep knowledge of the underlying political processes, the appropriate quantities of interest, and the substantive importance of any biases. Particularly with small-sample sizes, large standard errors, and/or highly nonlinear transformations, substantive scholars should consider application-specific simulations to assess the potential for bias. Closing the gap between methodological and substantive research requires mindful methodological work from both methodologists and substantive scholars.

The Concepts

As a motivating example, consider the log-linear model

$$\log(\text{income}_i) = \beta_{\text{cons}} + \beta_{\text{edu}}\text{education}_i + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$, education is measured in years, and income is measured in thousands of dollars. Assuming that the researcher uses the correct model, then least squares, which is also the ML estimator, provides the best unbiased estimator of the coefficients β_{cons} and β_{edu}. However, the researcher is not likely interested in log(income), but in income itself. In particular, she might want to estimate the median income among those with 20 years of education med(income | education = 20) $= e^{\beta_{\text{cons}}+20\beta_{\text{edu}}}$. Because med[$\log(y)$] = $\log$[med(y)] for a random variable y, one might guess that unbiased estimates of β_{cons} and β_{edu} lead to unbiased estimates of med(income | education = 20), but that is not the case. If we suppose that $N = 10$, $\beta_{\text{cons}} = 2.5$, $\beta_{\text{edu}} = 0.1$, $\sigma^2 = 1$, and education takes on integers roughly uniformly from 10 to 20, then $\tau(\beta_{\text{cons}}, \beta_{\text{edu}}) = e^{\beta_{\text{cons}}+20\beta_{\text{edu}}} \approx \$90k$. A simple Monte Carlo simulation, though, shows that although $\hat{\beta}_{\text{cons}}$ and $\hat{\beta}_{\text{edu}}$ are unbiased, the estimate of med(income | education = 20) is strongly biased upward, so that $E[\tau(\hat{\beta}_{\text{cons}}, \hat{\beta}_{\text{edu}})] = E(e^{\hat{\beta}_{\text{cons}}+20\hat{\beta}_{\text{edu}}}) \approx \$106k$.

A similar, but conceptually distinct issue arises when researchers want to calculate the *mean* from a log-linear model $\log(y) = X\beta + \epsilon$. Many textbooks highlight that $E[\log(y|X)] \neq \log[E(y|X)]$, so that $E(y|X_i) \neq e^{X_i\beta}$ (e.g., Wooldridge 2013, pp. 212-215). This inequality follows from a transformation of the random component of the model (i.e., ϵ_i). Even if the model coefficients β are *known*, then this inequality holds. But researchers can easily avoid this issue by using the correct transformation $E(y|X_i) = e^{X_i\beta + \frac{\sigma^2}{2}}$. However, the bias that interests me flows from a transformation of the *model coefficients*—even if the researcher uses the correct transformation, then $\hat{\tau}$ is biased.

But how does a simple transformation of unbiased coefficient estimates induce a large bias in the estimate of the quantity of interest? We usually think about bias as occurring in the model coefficients β, so that

$$\text{coefficient bias} = E(\hat{\beta}) - \beta.$$

But substantive researchers care mostly about bias in the quantities of interest. For convenience, I refer to the bias in the quantities of interest as τ-bias, so that

$$\tau\text{-bias} = E[\tau(\hat{\beta})] - \tau(\beta).$$

τ-bias is more complex and subtle than biases in the coefficients. It can be rewritten and decomposed into two components: transformation-induced τ-bias and coefficient-induced τ-bias, so that

$$\text{total } \tau\text{-bias} = \underbrace{E[\tau(\hat{\beta})] - \tau[E(\hat{\beta})]}_{\text{transformation-induced}} + \overbrace{\tau[E(\hat{\beta})] - \tau(\beta)}^{\text{coefficient-induced}}.$$

Any bias in the coefficients passes through to the quantities of interest in the sense that, if the coefficient estimates are biased, then the transformation of the true coefficient is not equal to the transformation of the average coefficient estimate, so that

$$\text{coefficient-induced } \tau\text{-bias} = \tau[E(\hat{\beta})] - \tau(\beta).$$

But the transformation *itself* introduces bias as well, so that

$$\text{transformation-induced } \tau\text{-bias} = E[\tau(\hat{\beta})] - \tau[E(\hat{\beta})].$$

Transformation-induced bias occurs because, in general, $h[E(y)] \neq E[h(y)]$ for an arbitrary random variable y and function h.

Little methodology research explicitly recognizes this transformation-induced τ-bias and less fully appreciates its practical importance. Both methodologists and substantive researchers must become more conscientious of transformation-induced τ-bias, which can be much larger than coefficient bias and disappear more slowly as the sample size increases.

A Characterization

But how can we characterize the direction of this bias? For strictly convex and strictly concave transformations, Jensen's inequality enables a straightforward characterization of the direction of the transformation-induced τ-bias. This characterization also provides the key intuition for more complicated transformations.

THEOREM 1. *Suppose a nondegenerate estimator $\hat{\beta}$. Then any strictly convex (concave) τ creates upward (downward) transformation-induced τ-bias.*

PROOF. The proof follows directly from Jensen's inequality. Suppose that the nondegenerate sampling distribution of $\hat{\beta}$ is given by $S_\beta(b)$ so that $\hat{\beta} \sim S_\beta(b)$. Then $E(\hat{\beta}) = \int_B b S_\beta(b) db$ and $E[\tau(\hat{\beta})] = \int_B \tau(b) S_\beta(b) db$. Suppose first that τ is convex. By Jensen's inequality, $\int_B \tau(b) S_\beta(b) db > \tau \left[\int_B b S_\beta(b) db \right]$, which implies that $E[\tau(\hat{\beta})] > \tau[E(\hat{\beta})]$. Because $E[\tau(\hat{\beta})] - \tau[E(\hat{\beta})] > 0$, the transformation-induced τ-bias is upward. By similar argument, one can show that for any strictly *concave* τ, $E[\tau(\hat{\beta})] - \tau[E(\hat{\beta})] > 0$ and that the transformation-induced τ-bias is downward. ∎

In general, researchers do not restrict themselves to a strictly convex or strictly concave τ. This situation is much more difficult to characterize generally because $\tau(b)$ might contain a mixture of convex and concave regions. For example, typical transformations of logistic regression coefficients, such as predicted probabilities, first and second differences, marginal effects, and risk ratios, all have both convex regions and concave regions. Making matters even more difficult, at any particular point b, the multivariate function τ might be convex in one direction and concave in another. In general, though, the direction of the bias depends on the *location* of the sampling distribution. But the intuition from Theorem 1 is clear. If most of the sampling distribution is located in a mostly concave region, then the bias will be downward. If most of the sampling distribution is located in a mostly convex region, then the bias will be upward.[1]

An Approximation

While Theorem 1 cultivates an intuition about *direction* of the bias, how can we assess the *magnitude* of the transformation-induced τ-bias? To approximate the magnitude, I use a second-order Taylor expansion. First, notice that $E[\tau(\hat{\beta})] = E[\tau(E[\hat{\beta}] + (\hat{\beta} - E[\hat{\beta}]))]$. Now approximate the term inside the right-hand expectation with a second-order Taylor expansion, so that

$$E[\tau(\hat{\beta})] \approx E\left[\tau[E(\hat{\beta})] + \sum_{r=1}^{k+1} \frac{\partial \tau[E(\hat{\beta})]}{\partial \beta_r}[\hat{\beta}_r - E(\hat{\beta}_r)] + \frac{1}{2} \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} \overbrace{\frac{\partial^2 \tau[E(\hat{\beta})]}{\partial \beta_r \beta_s}}^{\text{Hessian}=H_{rs}} \overbrace{[\hat{\beta}_r - E(\hat{\beta}_r)][\hat{\beta}_s - E(\hat{\beta}_s)]}^{\text{Cov}(\hat{\beta}_r, \hat{\beta}_s)=\Sigma_{rs}} \right].$$

Taking the expectation of the right-hand side eliminates the middle term and allows expressing the final term as a function of the variance of the sampling distribution, so that

$$E[\tau(\hat{\beta})] \approx \tau[E(\hat{\beta})] + \frac{1}{2} \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} H_{rs} \Sigma_{rs},$$

where H represents the Hessian matrix of second derivatives of τ at the point $E(\hat{\beta})$ and, conveniently, Σ represents the covariance matrix of the sampling distribution. Rearranging gives an approximation to the magnitude of the transformation-induced τ-bias, so that

$$\text{transformation-induced } \tau\text{-bias} = E[\tau(\hat{\beta})] - \tau[E(\hat{\beta})] \approx \frac{1}{2} \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} H_{rs} \Sigma_{rs}. \tag{1}$$

---

1 One might wonder about the relevance of these ideas to Bayesian analyses. Indeed, the researcher can usually use MCMC to sample directly from posterior of the model coefficients and, by simple extension, sample the quantity of interest from the posterior distribution. But if the researcher uses the posterior mode as the point estimate, then the identical logic applies. For an alternative point estimate (e.g., posterior mean), the invariance principle no longer holds, so the argument breaks down (i.e., the point estimate of $\tau(\beta)$ is not longer $\tau(\hat{\beta})$). However, regardless of the point estimate the researcher uses, a Bayesian approach does not guarantee an unbiased quantity of interest.

If H is constant then the approximation is exact. If $\hat{\beta}$ is unbiased, then $\tau[E(\hat{\beta})]$ can be replaced with $\tau(\beta)$, so that Equation (1) represents both transformation-induced and the total τ-bias.

Equation (1) does not depend on a strictly convex or concave transformation. As long as τ is not highly nonlinear (e.g., $|\frac{\partial^3 \tau}{\partial \beta_r \partial \beta_s \partial \beta_t}| \approx 0$), then Equation (1) provides a reasonable estimate of the direction and magnitude of the bias.

Equation (1) quantifies two intuitions. First, the amount of bias depends on the standard error or sample size. As the sample size grows large, Σ_{rs} shrinks to zero, which drives the bias to zero as well. This matches the previous observation that $\tau(\hat{\beta})$ is a consistent estimator of $\tau(\beta)$. Secondly, the amount of bias depends on the curvature in τ. If τ is nearly linear so that $H \approx 0$, then the transformation introduces minimal bias. On the other hand, more curvature, so that $H >> 0$, leads to a large bias.

Two Monte Carlo Simulations

But does this bias matter in practice? The following two Monte Carlo studies illustrate the importance of accounting for transformation-induced bias when evaluating estimators.[2] Approximately unbiased coefficients are not enough—one must assess the bias in the quantities of interest as well.

A hypothetical model

Many substantive researchers realize that logistic regression estimates are biased away from zero in small samples and use "rules of thumb" to judge whether asymptotic properties, such as asymptotic unbiasedness, approximately apply to a finite sample. When nonevents outnumber events, one such rule of thumb requires ten events per explanatory variable (Peduzzi *et al.* 1996). I show that this rule works quite well choosing a sample size that yields approximately unbiased coefficients, but severely underestimates the sample size needed for approximately unbiased estimates of the marginal effects.

For simplicity, I focus on the model $\Pr(y) = \text{logit}^{-1}(\beta_{\text{cons}} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6)$, where y indicates whether or not each observation experiences an event and the x_j represent fixed explanatory variables that I create by simulating from independent, standard normal distributions. For this simulation, I set $\beta_{\text{cons}} = -1$ and $\beta_j = 0.15$ for $j \in \{1, \ldots, 6\}$. I assume that "approximately unbiased" means a bias of less than three percent, where

$$\text{percent bias} = 100 \times \frac{E[\tau(\hat{\beta})] - \tau(\beta)}{\tau(\beta)}. \tag{2}$$

I vary number of observations N from 100 to 3,000, and, for each sample size, I simulate 100,000 data sets, use each data set to estimate the coefficients, and use the estimated coefficients to calculate the marginal effects. I use these 100,000 estimates to calculate the percent bias given by Equation (2).

Figure 1 shows the bias in the coefficients as the sample size increases. The left panel shows the bias in $\hat{\beta}_{\text{cons}}$ and the right panel shows the bias in $\hat{\beta}_1$. For $N = 100$, $\hat{\beta}_{\text{cons}}$ and $\hat{\beta}_1$ are biased away from zero by about ten percent. However, this bias drops to about three percent for $N = 250$ and nearly disappears for $N = 3,000$. The rule of thumb works well; the bias is negligible for about $N = 219$.

Figure 2 shows the bias in the estimates of the marginal effects as the sample size increases. The left panel shows the total bias, the middle panel shows the coefficient-induced bias, and the right panel shows the transformation-induced bias. Since the marginal effect of x_1 varies with x_1 itself, I plot the estimates for a range of values of x_1.

---

2 All computer code necessary to reproduce these simulations as well as the data analysis below can be found at github.com/carlislerainey/transformation-induced-bias and dx.doi.org/10.7910/DVN/CYXFB8 (Rainey 2017).

Figure 1. This figure shows the percent bias for the intercept and coefficient for x_1. The rule of thumb requiring ten events per explanatory variable suggests a minimum sample size of about 219. For samples larger than about 250, the bias falls below three percent and it nearly disappears as the sample size approach 3,000.
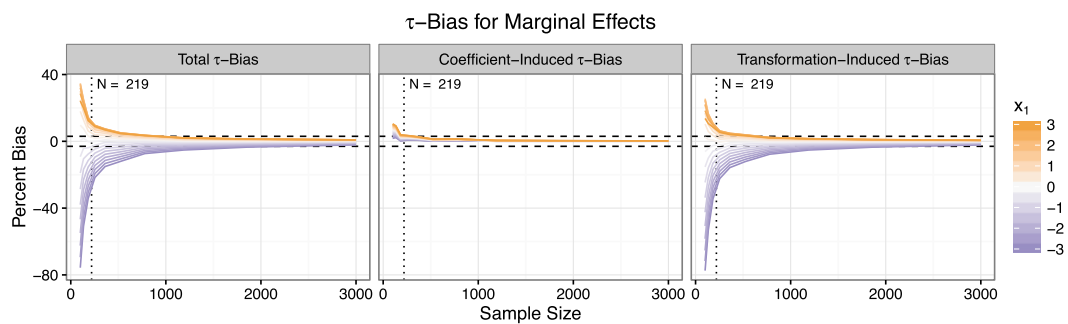

Figure 2. This figure shows the total, coefficient-induced, and transformation-induced τ-bias for the marginal effects. The rule of thumb requiring ten events per explanatory variable suggests a minimum sample size of about 219. However, the bias falls well outside the three percent threshold for this suggested sample size. The estimates fall within the three percent threshold only for sample sizes nearing 3,000—more than ten times the rule of thumb that works well for the coefficients. Also notice that while the coefficient-induced bias receives the most attention from methodologists, the transformation-induced bias is *much* larger.

Two features stand out. First, small-sample bias is much larger for the marginal effects than for the coefficients. For $N = 100$, the estimate of the marginal effect is biased by about -75% for $x_1 = -3$, -50% for $x_1 = -2$, and -25% for $x_1 = -1$. Second, the small-sample bias in the estimates of the marginal effects descends to zero more slowly than the coefficient estimates. While the coefficient estimates are approximately unbiased for about $N = 250$, the estimates of the marginal effects retain substantial bias. Indeed, the bias in the estimates of the marginal effects drops below the 3% threshold at about $N = 3,000$—more than ten times the rule of thumb that works well for the coefficients.

An actual model

To further highlight the practical implications of transformation-induced τ-bias, I use the explanatory variables and coefficients reported for Model 1 in Table 2 of Lacina (2006, p. 286) to conduct a second simulation. Lacina uses a normal-linear regression model with a log-transformed outcome variable to assess several hypotheses about the causes of the number of battle deaths in civil conflicts. Using her reported coefficients as the true model parameters and her explanatory variables the predictors (105 complete observations; 9 explanatory variables), I repeatedly (1) simulate a new outcome variable, (2) re-estimate the regression model, and (3)
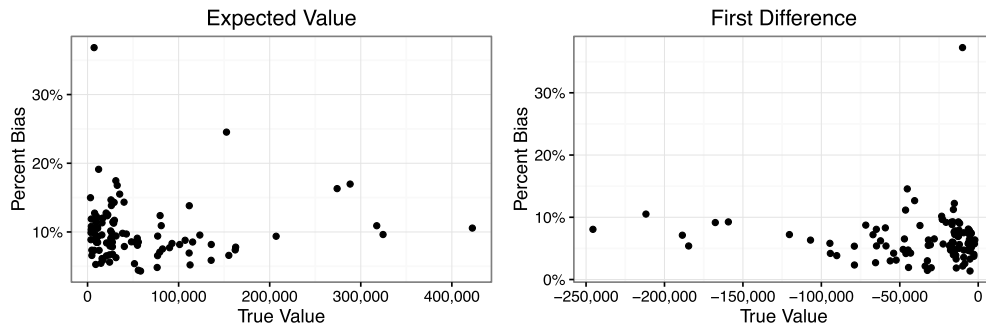
Figure 3. This figure shows the transformation-induced τ-bias for two quantities of interest. Each point represents a single observation from Lacina's (2006) data set. For each observation, I calculate the transformation-induced τ-bias in the expected value—the expected number of battle deaths—and in the first difference—the change in the expected number of deaths if each case was changed from a nondemocracy to a democracy.

calculate the quantities of interest. For each set of estimates $\hat{\beta}$ and $\hat{\sigma}^2$, I calculate two quantities of interest:

(1) The expected number of battle deaths $E(\text{deaths}_i \mid X_i) = e^{X_i\hat{\beta}+\frac{\hat{\sigma}^2}{2}}$ for each observed case X_i.

(2) The first difference (i.e., the change in the expected number of battle deaths) $\Delta(\text{deaths}_i \mid X_i^D, X_i^{\sim D}) = E(\text{deaths}_i \mid X_i^D) - E(\text{deaths}_i \mid X_i^{\sim D}) = e^{X_i^D\hat{\beta}+\frac{\hat{\sigma}^2}{2}} - e^{X_i^{\sim D}\hat{\beta}+\frac{\hat{\sigma}^2}{2}}$ if each observed case X_i were changed from a nondemocracy $X_i^{\sim D}$ to a democracy X_i^D.

Because the estimates of β and σ^2 are unbiased, there is no coefficient-induced τ-bias. Indeed, the least squares estimate of β is the best unbiased estimator under the assumed normal-linear model. However, this ideal small-sample property does not apply to the quantities of interest. Figure 3 summarizes these simulations and demonstrates that transformation-induced τ-bias can create considerable bias in the quantities of interest even when coefficient estimates have optimal properties.

The left panel of Figure 3 shows the true value and percent bias for the expected number of battle deaths $E(\text{deaths}) = e^{X\hat{\beta}+\frac{\hat{\sigma}^2}{2}}$. Because the estimates $\hat{\beta}$ and $\hat{\sigma}^2$ are unbiased, there is no coefficient-induced τ-bias in the expected value. However, there is a substantial upward bias in the expected value due to transformation-induced τ-bias. The upward bias in the expected value ranges from 4% (307 deaths) to 37% (51,120 deaths). The average upward bias is 10%—these are not trivial biases.

The right panel of Figure 3 shows that these upward biases do not cancel for the first difference. Indeed, transformation-induced τ-bias leads to an overly optimistic estimate of the effect of democracy. Lacina correctly notes that "democracy is associated with fewer battle deaths" (p. 287), but transformation-induced τ-bias might lead researchers to over-estimate this pacifying effect by up to 10% (and about 35% in one extreme case) and 6% on average.

The Implications

Quantities of interest do not inherit the small-sample properties of the coefficient estimates. This fact has important implications for how we evaluate the small-sample properties of estimators.

First, τ-bias has important implications for the sample sizes that methodologists recommend to substantive researchers. Methodologists usually parameterize models so that the coefficients lie in an unbounded space. This allows the coefficient estimates to rapidly approach their asymptotic distribution, which ensures the estimates have acceptable small-sample properties. Substantive researchers, though, usually transform these coefficient estimates into a quantity of interest, which, because it often lies in a bounded space, might approach its asymptotic distribution

more slowly. As a result, substantive researchers might need much larger sample sizes than methodologists usually recommend. Methodologists must remain conscientious of the quantities of interest to substantive researchers and assess the performance of their estimators in terms of these quantities. Unfortunately, it remains impossible or difficult to assess the bias in general or for a wide range of sample sizes, quantities of interest, or parameter values.

But substantive researchers must also remain aware of the potential to introduce bias into estimates by transforming coefficient estimates. Fortunately, substantive researchers can use Monte Carlo simulations to quickly assess the potential for bias in a specific substantive context in which researchers know the sample size, quantity of interest, and likely parameter values.

Secondly, τ-bias has important implications for the bias-variance tradeoff in choosing an estimator. Methodologists usually recognize a tradeoff between bias and variance in estimating parameters. Actions intended to remove bias might increase variance and vice versa. However, the approximation to the transformation-induced τ-bias given in Equation (1) points out an important result. Greater variance in the coefficient estimates might lead to increased bias in the quantities of interest. This implies that if an estimator is essentially unbiased, then greater efficiency translates to reduced bias in the quantities of interest. Similarly, small reductions in bias at the expense of a large increase in variance might lead to greater bias in the quantities of interest. For example, refinements of the usual logit model intended to *reduce* bias in the coefficients, such as heteroskedastic probit or scobit, might actually *increase* bias in the quantities of interest. Methodologists must be aware of this tradeoff when recommending more complex estimators to substantive researchers and comparing alternative estimators.

Methodologists cannot ignore transformation-induced bias. Substantive researchers must not assume that sample size recommendations remain valid for any quantity of interest. Nearly unbiased estimates of coefficients are not enough. We must remain thoughtful about our quantities of interest and calibrate our tools for these quantities.

References

Casella, George, and Roger L. Berger. 2002. *Statistical inference*, 2nd edn. Pacific Grove, CA: Duxbury.

King, Gary. 1998. *Unifying political methodology: The likelihood theory of statistical inference*. Ann Arbor: Michigan University Press.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science* 44(2):341–355.

Lacina, Bethany. 2006. Explaining the severity of civil wars. *Journal of Conflict Resolution* 50(2):276–289.

Long, J. Scott. 1997. Regression models for categorical and limited dependent variables. In *Advanced quantitative techniques in the social sciences*, Thousand Oaks, CA: Sage.

Nagler, Jonathan. 1994. Scobit: An alternative estimator to logit and probit. *American Journal of Political Science* 38(1):230–255.

Nieman, Mark David. 2015. Statistical analysis of strategic interaction with unobserved player actions: Introducing a strategic probit with partial observability. *Political Analysis* 23(3):429–448.

Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49(12):1373–1379.

Rainey, Carlisle. 2017. Replication Data for: Transformation-Induced Bias. doi:10.7910/DVN/CYXFB8. Harvard Dataverse, V1, UNF:6:XDVZ8wD2BMxScpCoFcCLYg==.

Wooldridge, Jeffrey M. 2013. *Introductory econometrics: A modern approach*, 5th edn. Mason, OH: South-Western Cengage Learning.